The Reconstruction of Financial Signals using FastICA for Systemic Risk

Kuan-Heng Chen
Financial Engineering Program
Stevens Institute of Technology
Hoboken, NJ, USA
kchen3@stevens.edu

Khaldoun Khashanah
Financial Engineering Program
Stevens Institute of Technology
Hoboken, NJ, USA
kkhashan@stevens.edu

***Abstract*— Independent component analysis (ICA) is a statistical method for transforming multidimensional observed signals into components, which are statistically independent from each other, which is a case of redundancy reduction. In this paper, we implement FastICA proposed by Hyvarinen and Oja to investigate the relationship between systemic risk and ICA in the US financial market. We propose a systemic risk indicator based on observing the redundancy level of signals in running ten variables including ten S&P 500 sector indices. We find that not only the redundancy level of signals becomes larger during a crisis than during a normal period, but also the financial system becomes more vulnerable when the redundancy level grows up.**

I. INTRODUCTION

The recent financial subprime crisis showed that financial innovations that increased the complexity and interconnectedness of aspects of the financial system increased as a consequence of system's vulnerability to financial shocks. In response to the vulnerability it revealed, governments around the globe are acting to improve financial stability and reduce the risks posed by a highly interconnected financial system, which is systemic risk. As a result, it is important to develop an indicator for measuring systemic risk as well as the financial stability.

The Commodity Futures Trading Commission[1] (CFTC) defined systemic risk as "*The risk that a default by one market participant will have repercussions on other participants due to the interlocking nature of financial markets.*" The Financial Stability Board [2] stated that "*systemic risk can arise through direct and indirect interlinkages between the components of the financial system so that individual failure or malfunction has repercussions around the financial system.*" This is also called the domino effect, and the term domino effect is sometimes used when interconnectedness causes the failure of one entity to result in the failure of others. Although there are several different definitions of systemic risk, a commonality in the various definitions of systemic risk is that a trigger event causes financial systems to become unstable because of high interconnectedness.

So, what have we learned from recent research on interconnectedness and systemic risk? We have seen lots of research studies on how interconnectedness can be a driver of systemic risk. The aggregated risk may do little harm or may even be irrelevant in normal periods, but they can be devastating to a financial system during a crisis. As a result, systemic risk measure is a "Too Interconnected to Fail" assessment. On the other hand, systemic risk is measured by how correlated the entire financial system exists.

Billio *et al.* [3][4] and Kritzman *et al.* [10] used principal component analysis (PCA) to gauge the degree of the financial system connectedness. They claimed that the markets are much more fragile during a tight interconnection than during loose interconnections. Although many researchers use PCA to identify the key principal components from a large number of mixed signals as well as to explain the relationship between the interconnectedness and systemic risk, PCA does not use any high-order statistics and ignores features such as data clustering [8]. As a result, PCA and factor analysis (FA) are not suitable for modeling multivariate financial time series because most of the financial returns are non-Gaussian distributions; in addition, the two random variables that are uncorrelated do not imply they are independent. Since ICA takes into account the whole dependence structure of the variables, ICA is a preferable measure in analyzing multivariate financial time series. Chen *et al.* [5] showed that ICA is a complementary tool compared to PCA, allowing the capture of high-order properties of financial time series.

Independent component analysis was originally developed to deal with problems that are closely related to the cocktail-party problem. Comon [6] explained how to obtain a more general ICA formulation, which does not need to assume an underlying data distribution, and a large number of ICA algorithms have been developed. One of the best methods is the FastICA algorithm proposed by Hyvarinen and Oja [8]. In the FastICA algorithm, the mutual information is used as the criterion to estimate independent components, minimizing the mutual information between the components corresponds to maximizing their negentropy. The algorithm is quite simple, converges fast and reliably.

Until recent years, ICA applications have been used in the financial field [1][5], and people usually use ICA to decompose multivariate time series into statistically independent time series. Although much research has investigated the relationship between dependence and systemic risk, there is no research in measuring systemic risk based on ICA.

1. Definition of "Systemic Risk" in the CFTC GLOSSARY. http://www.cftc.gov/ConsumerProtection/EducationCenter/CFTCGlossary/index.htm#S

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.130

ICA linearly transforms the original mixed signals to a set of independent signals, whose dependence is removed. Since the ICA removes the dependence from the original signals, we explore the relationship between redundancy (interconnectedness) and systemic risk by observing how much redundancy has been removed over time. The main finding in this paper indicates that the higher the degree of system interconnectedness, the more vulnerable the financial stability.

This paper has four sections. The first section briefly introduces existing research regarding systemic risk and independent component analysis in the financial field. The second section describes the methodology of independent components analysis. The third section presents the data and outlines how we apply ICA measure to evaluate systemic risk. The fourth section concludes our findings.

II. METHODOLOGY

A. Independent Component Analysis (ICA)

Independent component analysis (ICA) is a multivariate analysis technique which separates or recovers mutually independent but unknown sources from the linear mixtures observations [6][8][9]. Let X be an observed *m* vector. The ICA model for X is written as

$$X = AS \tag{1}$$

where A is called a mixing matrix, S is an *n* vector of latent independent components and $n \leq m$. The formula is basically the same as classical factor analysis, with the crucial difference that the independent components S are assumed to be non-Gaussian. The main goal of ICA is to estimate the mixing matrix A by maximizing independence among the components of S. Since the independence is measured by non-normality, the estimation is implemented by finding the demixing matrix $\widehat{W}$ such that the components of $\hat{S} = \widehat{W}X$ have maximum non-Gaussianity. A typical measure of non-normality is excess kurtosis, and the excess kurtosis is zero for a normal distribution and non-zero for most non-normal distributions. Although the concept of ICA using kurtosis is simple, it can be very sensitive to outliers. Therefore, kurtosis is not a robust measure of non-Gaussianity.

Entropy is the essential concept of information theory, and it interprets the degree of information that the observation of the variable gives. Entropy H is defined for a discrete random variable Y as

$$H(Y) = -\sum_i P(Y = y_i) \log P(Y = y_i) \tag{2}$$

Negentropy is based on the quantity of differential entropy. To obtain a measure of non-Gaussianity that is zero for a Gaussian variable and always non-negative, negentropy J is defined as

$$J(Y) = H(Y_{gauss}) - H(Y) \tag{3}$$

where Y_{gauss} is a Gaussian random variable of the same covariance matrix as Y. In addition, negentropy has the property that it is invariant for invertible linear transformations [6]. Hyvarinen and Oja [8] developed a new approximation for negentropy based on the maximum-entropy principle, and it can be expressed as

$$J(Y) \propto \{E[G(Y)] - E[G(v)]\}^2, v \sim N(0,1) \tag{4}$$

where G is a nonlinear and non-quadratic function, i.e. $G(x) = \log\cosh(x)$.

B. FastICA Algorithm

Hyvarinen and Oja [8] proposed a very efficient algorithm to estimate W by maximizing negentropy J in (4), called FastICA. It is assumed that the random vector x is whitened (sphered), uncorrelated and has variance of one. The fixed-point algorithm for p units is as follow:

1) Choose an initial random vector for the first unit w_1 of norm one
2) Let $w_i^+ = w_i - \mu \left\{ \frac{E[xg(w_i^T x) - \beta w_i]}{(E[g'(w_i^T x)] - \beta) I} \right\}$
3) Normalize $w_i^+ \leftarrow \frac{w_i^+}{\|w_i^+\|}$. If not converges, go back to step 2.
4) Decorrelate the outputs w_j after every iteration for unit $j = 1, \ldots, p$. Let $w_p = w_p - \sum_{j=1}^{p-1} w_p^T w_j w_j$
5) Back to step 2 to run the one-unit fixed-point algorithm unitl p units achieved.
6) $\hat{S} = \widehat{W}X$

where μ is a step size, g is the derivative of G, $\beta = E\{w^T x g(w^T x)\}$, and whitening can be accomplished by principal components analysis (PCA). Based on the Kuhn-Tucker condition and the Newton's method, Hyvarinen and Oja [8] showed that the estimations based on the FastICA algorithm are consistent.

C. Systemic Risk Indicator

At present, there is no better method available to automatically determine the optimum number of independent components. In our case, we determine the number of independent components to be ten because each industrial sector cannot be substituted. The centering observations X are ten vectors including the returns of ten S&P 500 sector indices, and the reconstructed signals are $\hat{X}$, generated by the estimated ten independent components $\hat{S}$ and the estimated mixing matrix $\hat{A}$. The concept of systemic fragility is the degree of system's susceptibility to structural breakdown, and our paper defines evaluating the degree of systemic fragility as a systemic risk indicator. On the other hand, an increase in systemic fragility as measured in this paper is an indication of elevated interdependency risk. The main difference between ICA and PCA is that ICA not only decorrelates the signals up to second-order statistics, but also reduces the dependency up to higher-order statistics. Therefore, using ICA to capture the dependency for constructing the systemic risk indicator is more reliable than using PCA. In addition, the potential for systemic risk exists not only because of the degree of the risk repercussion, but also because of the sheer size of an exposure to other participants. The larger size of boom tends to increase the probability of system breakdown. As a result, the dynamic systemic risk indicator (SRI) is composed by the weighted sum of the absolute value of the residual time series, and it can be written as

$$SRI_t = \sum_{i=1}^{10} w_{i,t} \cdot \left| X_{i,t} - \widehat{X_{i,t}} \right| \tag{5}$$

where w is the weight vectors calculated by the ratio of the sector's capitalization to the market's capitalization.

III. Empirical Results

A. Data Representation

Standard and Poor's corporation divides the 500 members from the S&P 500 index into ten different sector indices based on the Global Industrial Classification Standard (GICS) shown in table 1 below. Table 1 shows that it is reasonable to develop the systemic risk indicator by the weighted sum since the size of the information technology sector is approximately seven times larger than the size of the telecommunication services sector.

TABLE I. THE 10 S&P 500 SECTOR INDICES, SNAPPED BY MAR 28, 2014

Sector Name	Index Weight (%)
Information Technology	18.6
Financials	16.4
Health Care	13.3
Consumer Discretionary	12.1
Industrials	10.6
Energy	10.2
Consumer Staples	9.7
Materials	3.5
Utilities	3.1
Telecommunication Services	2.5

TABLE II. THE STATISTICS OF S&P 500 SECTOR INDICES AND THE S&P 500 INDEX

	Mean	Sigma	Skew	Kurt	ADF test	J-B test
S&P 500 Financials	0.02%	1.91%	-0.102	18.08	1	1
S&P 500 Information Technology	0.04%	1.82%	0.146	7.716	1	1
S&P 500 Consumer Discretionary	0.04%	1.37%	-0.112	9.924	1	1
S&P 500 Energy	0.03%	1.60%	-0.303	13.4	1	1
S&P 500 Health Care	0.04%	1.20%	-0.136	9.255	1	1
S&P 500 Industrials	0.03%	1.33%	-0.338	8.571	1	1
S&P 500 Utilities	0.01%	1.17%	-0.027	12.933	1	1
S&P 500 Consumer Staples	0.03%	0.97%	-0.128	11.349	1	1
S&P 500 Materials	0.02%	1.51%	-0.245	9.531	1	1
S&P 500 Telecommunication Services	0.01%	1.43%	0.056	9.458	1	1

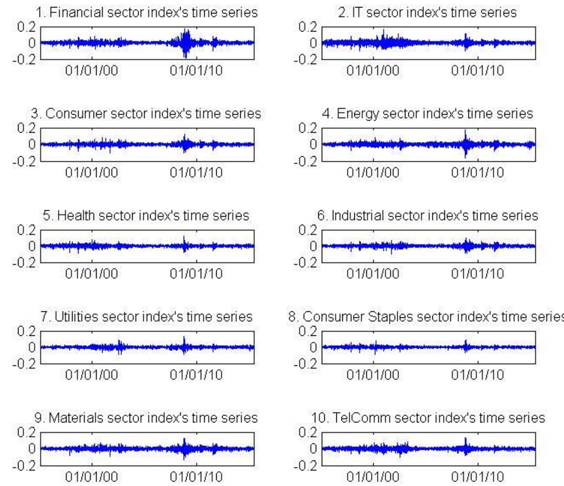

Fig. 1. The S&P 500 sector indices' returns time series.

The daily data is taken from Bloomberg from January 1, 1995 to June 26, 2015, covering 5156 observations shown in Fig. 1. Table 2 presents the summary statistics of the ten S&P 500 sector indices and S&P 500 index. As shown in table 2, the Augmented Dickey-Fuller test (ADF) shows that each series returns reject the null hypothesis of a unit root, indicating each series of the returns is stationary. The result is 1 if the test rejects the null hypothesis at the 5% significance level in our case. Meanwhile, the Jarque-Bera (J-B) test shows that each series returns reject the null hypothesis that the data comes from a normal distribution with an unknown mean and variance, indicating each series of the returns does not follow an normal distribution. The result is 1 if the test rejects the null hypothesis at the 5% significance level. In addition, the statistical analysis interprets that it is more appropriate to use ICA to measure systemic risk than use PCA because each distribution of the time series violates normality.

B. Empirical Results of Systemic Risk Indicator

In the extending rolling sample approach, the observations increase over time. For instance, the first value is calculated by the observations from day 1 to day 1000. The second value is calculated by the observations from day 1 to day 1001, and so on. As for the general purpose, we determine the contrast function to be $G(x) = \log\cosh(x)$. Fig. 2 interprets the estimates of the original source signals, and Fig. 3 provides the residual time series between the original signals and the replicated signals in the last time window.

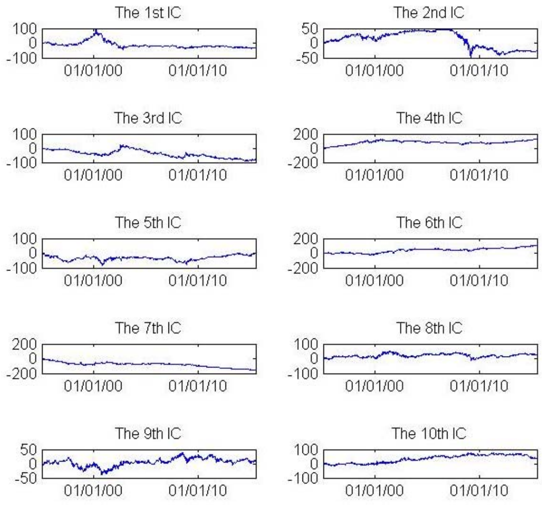

Fig. 2. The estimates of the original source signals.

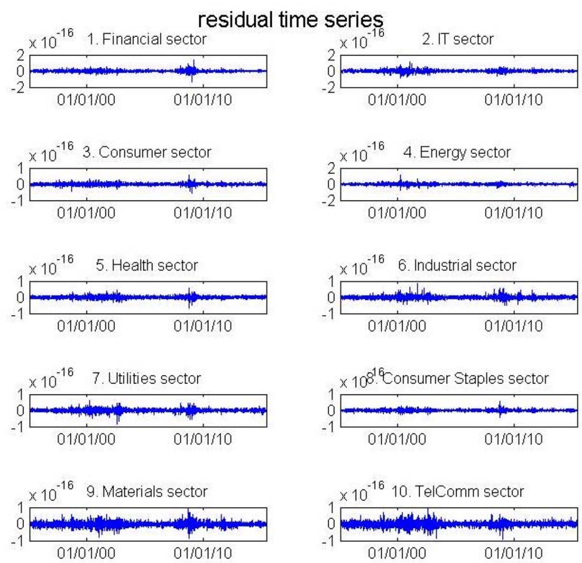

Fig. 3. The comparison between the original signals and the replicated signals in last time window.

Fig. 3 shows that the information technology sector was the major risk contribution to Dot-com bubble in 2000, and the financial sector dominated the subprime crisis in 2008. Here, we determine 10 PCs and 10 ICs. Table 3 shows that the correlation matrix in the last time window.

TABLE III. THE CORRELATION MATRIX

	Financials	Information Technology	Consumer Discretionary	Energy	Health Care	Industrials	Utilities	Consumer Staples	Materials	Telecommunication Services
Financials	1	0.5901	0.7830	0.5516	0.6081	0.7951	0.5117	0.6055	0.6780	0.5885
Information Technology	0.5901	1	0.7087	0.4275	0.4938	0.7058	0.3752	0.4240	0.5499	0.5700
Consumer Discretionary	0.7830	0.7087	1	0.5692	0.6588	0.8624	0.5275	0.6630	0.7373	0.6403
Energy	0.5516	0.4275	0.5692	1	0.5271	0.6352	0.5872	0.5269	0.6945	0.4638
Health Care	0.6081	0.4938	0.6588	0.5271	1	0.6757	0.5262	0.7248	0.5616	0.5280
Industrials	0.7951	0.70580	0.8624	0.6352	0.6757	1	0.5657	0.6758	0.8023	0.6253
Utilities	0.5117	0.3752	0.5275	0.5872	0.5262	0.5657	1	0.5630	0.5309	0.4983
Consumer Staples	0.6055	0.4240	0.6630	0.5269	0.7248	0.6758	0.5630	1	0.5958	0.5390
Materials	0.6780	0.5499	0.7373	0.6945	0.5616	0.8023	0.5309	0.5958	1	0.5224
Telecommunication Services	0.5885	0.5700	0.6403	0.4638	0.5280	0.6253	0.4983	0.5389	0.5224	1

By observing the redundancy level, Fig. 4 reports that the financial market becomes highly interconnected during financial crises and relatively uncoupled during economic expansions. On the other hand, it means that the financial market is much more fragile during a tight interconnection than during loose interconnections. Fig. 4 represents that the global stock market crash was caused by the Asian economic crisis in 1997, and the Russian government devalued the Ruble and defaulted on its debt in 1998. Most recently, the collapse of a technology bubble existed in 2000, and the financial crisis was triggered by the subprime mortgage in 2008. In our study, we set the threshold of the systemic risk indicator be the 99 percentile as the red line shown in Fig. 4 (b).

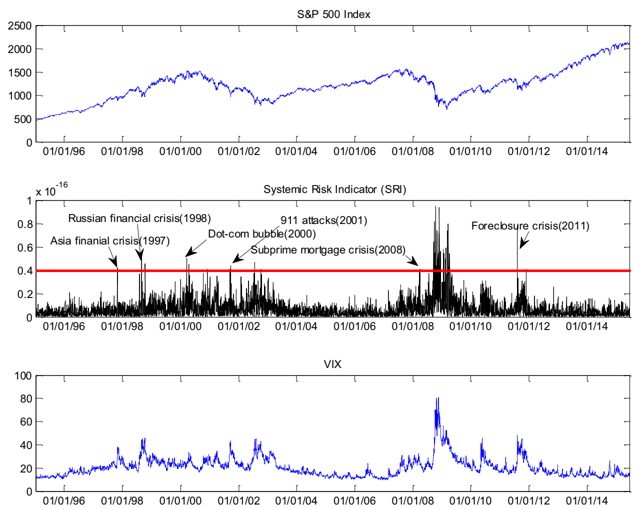

Fig. 4. (a) S&P 500 Index; (b) Systemic Risk Indicator; (c) VIX.

IV. CONCLUSION

ICA is a distribution-free and general-purpose statistical technique in which observed data are linearly transformed into components that are statistically independent. Although applications of ICA can be found in many different fields, this paper is the first to explore systemic risk based on ICA. In this paper, the technique has been applied to an econometric framework with the objective of estimating the redundancy level that reveals a driving mechanism of the financial stability. With the technique of removing dependence among multivariate time series based on ICA, the main findings are that the US financial system not only becomes more interconnected during economic recession than during economic expansion, but also becomes more vulnerable when the redundancy level grows up. In additional, using our methodology to measure systemic risk in the US financial market is a simple, efficient and useful method.

REFERENCES

[1] Back, Andrew D., and Andreas S. Weigend. "A first application of independent component analysis to extracting structure from stock returns." *International journal of neural systems* 8, no. 04 (1997): 473-484.

[2] Board, Financial Stability. "Guidance to assess the systemic importance of financial institutions, markets and instruments: initial considerations." *Report to G20 Finance Ministers and Governors* (2009).

[3] Billio, Monica, Mila Getmansky, Andrew W. Lo, and Loriana Pelizzon.*Econometric measures of systemic risk in the finance and insurance sectors*. No. w16223. National Bureau of Economic Research, 2010.

[4] Billio, Monica, Mila Getmansky, Andrew W. Lo, and Loriana Pelizzon. "Econometric measures of connectedness and systemic risk in the finance and insurance sectors." *Journal of Financial Economics* 104, no. 3 (2012): 535-559.

[5] Chen, Ying, Wolfgang Härdle, and Vladimir Spokoiny. "Portfolio value at risk based on independent component analysis." *Journal of Computational and Applied Mathematics* 205, no. 1 (2007): 594-607.

[6] Comon, Pierre. "Independent component analysis, a new concept?." *Signal processing* 36, no. 3 (1994): 287-314.

[7] Hyvarinen, Aapo. "Fast and robust fixed-point algorithms for independent component analysis." *Neural Networks, IEEE Transactions on* 10, no. 3 (1999): 626-634.

[8] Hyvärinen, Aapo, and Erkki Oja. "Independent component analysis: algorithms and applications." *Neural networks* 13, no. 4 (2000): 411-430.

[9] Hyvärinen, Aapo, Juha Karhunen, and Erkki Oja. *Independent component analysis*. Vol. 46. John Wiley & Sons, 2004.

[10] Kritzman, Mark, Yuanzhen Li, Sebastien Page, and Roberto Rigobon. "Principal components as a measure of systemic risk." *The Journal of Portfolio Management* 37, no. 4 (2011): 112-126.

[11] Kiviluoto, Kimmo, and Erkki Oja. "Independent Component Analysis for Parallel Financial Time Series." In *ICONIP*, vol. 2, pp. 895-898. 1998.

[12] Oja, Erkki, Kimmo Kiviluoto, and Simona Malaroiu. "Independent component analysis for financial time series." In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pp. 111-116. IEEE, 2000.